REMARKS

I. Status Summary

Claims 1-23 are pending in the present application. Claims 1 and 11 have been amended. Therefore, upon entry of this Amendment, Claims 1-23 will be pending. No new matter has been introduced by the present amendment. Reconsideration of the application as amended and based on the arguments set forth hereinbelow is respectfully requested.

The Examiner indicated that the article titled "Combined Acoustic Echo Control and Noise Reduction for Hand-Free Telephony", S. Gustafsson et al. (Signal Processing 64 (1998), was not received. Applicant has included herewith a copy of the article.

II. Specification

The abstract of the disclosure is objected to because it contains "Fig. 3". (Official Action, page 2.) The phrase "Fig. 3" at the last line of the Abstract has been deleted. Therefore, applicant respectfully submits that the objection to the abstract of the disclosure should be withdrawn.

III. Claim Rejections Under 35 U.S.C. § 102

Claims 1, 2, 4, 5, 11, 12, 15-20, 22, and 23 stand rejected under 35 U.S.C. § 102(e) as being anticipated by U.S. Patent No. 5,933,495 to Oh (hereinafter, "Oh"). This rejection is respectfully traversed.

Regarding Claim 1, the Examiner contended that Oh teaches a device at Figure 2 for subband noise suppression in telephone devices using a subband adaptive filter

216. (Official Action, page 3.) The Examiner stated that Oh also teaches a control circuit for adjusting filter coefficients operating in the subband and synthesis filter **234** transforms the subband reduced-noise signal into a full-band signal at Figures 2 and 3, and column 4, lines 9-67 of Oh. (Official Action, page 3.)

Upon careful consideration and review of Oh, applicant respectfully submits that Oh does not disclose each and every element of the presently claimed subject matter and therefore does not anticipate the presently claimed subject matter. Claim 1 recites a device for suppressing noise in telephone equipment. Further, Claim 1 recites an additional filter with a short propagation time being arranged in the transmission path of the telephone equipment. Claim 1 has been amended to recite that the additional filter includes adjustable coefficients and a control circuit for adjusting the coefficients. The additional filter operates in the full band while the control circuit for adjusting the coefficients operates in the subband. Applicant respectfully submits that Oh does not disclose these features required by amended Claim 1.

According to the Examiner, Figure 2 of Oh teaches adaptive filter **216** including a control circuit for adjusting the filter coefficients operating in the subband and synthesis filter **234** for transforming the subband reduced-noise signal into a full-band signal. (Official Action, page 3.) Referring to Figure 2 of Oh, adaptive filter **216** operates in a subband. Adaptive filter **216** operates similarly to adaptive filter **116**, which is described as operating in the subband. (Oh, column 1, lines 59-63, and column 4, lines 22-25.) In addition, Oh teaches that the coefficients of adaptive filter **116** are adjusted to provide acoustic echo cancellation. (Oh, column 2, lines 2-4.) The coefficients are provided via line **228** to adaptive filter **216** in an acoustic echo

canceller block **210**. (Oh, Figure 2.) Thus, Oh teaches that filtering occurs in the subband. In marked contrast, Claim 1 recites that the additional filter operates in the full band. For these reasons, Oh does not teach each and every feature of Claim 1 and, thus, cannot anticipate the claim.

Claims 2, 4, and 5 depend from Claim 1. Therefore, claims 2, 4, and 5 include the features of Claim 1. Thus, the comments presented below relating to amended Claim 1 apply equally to claims 2, 4, and 5. For the same reasons provided for Claim 1, it is respectfully submitted that Oh does not anticipate Claims 2, 4, and 5.

The Examiner stated that Claim 11 is similar to Claim 1 and rejected for the same reasons. Claim 11 has been amended to place the claim in better method claim format. Claim 11 recites a method for noise suppression in the telephone equipment. Claim 11 has been amended to recite a step for filtering the transmitted signal from the telephone equipment with a short propagation time. In addition, Claim 11 has been amended to recite a step for controlling the filtering of step (a) with adjustable coefficients. Further, Claim 11 recites that the filtering is carried out in the full band, while the determination of the coefficients is carried out in the subband. Applicant respectfully submits that Oh does not disclose these features recited by amended Claim 11.

As previously stated, Oh teaches adaptive filter **216** operating in a subband within block **210**. Further, Oh teaches that the coefficients of adaptive filter are provided via line **228** to adaptive filter **216** in block **210**. Oh also teaches that block **210** operates in the subband, not the fullband. Thus, Oh teaches that filtering occurs in the subband. In marked contrast, Claim 11 recites that the filtering is carried out in

the full band while the determination of the coefficients is carried out in the subband. For these reasons, Oh does not teach each and every feature of Claim 11 and, thus, cannot anticipate the claim.

Claims 12, 15-20, 22, and 23 depend from Claim 11. Therefore, Claims 12-23 include the features of Claim 11. Thus, the comments presented below relating to amended Claim 11 apply equally to claims 12-23. For the same reasons provided for Claim 11, it is respectfully submitted that Oh does not anticipate Claims 12-23.

For the all of the reasons provided above, applicant respectfully requests that the rejections of Claims 1, 2, 4, 5, 11-12, 15-20, 22 and 23 under 35 U.S.C. §102(e) be withdrawn and the claims allowed at this time.

IV. Claim Rejections Under 35 U.S.C. § 103

Claims 6-9 and 21 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Oh, as applied to Claims 5 and 20, and further in view of U.S. Patent No. 5,757,937 to Itoh et al. (hereinafter, "Itoh"). In addition, Claims 3, 10, 13, and 14 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Oh as applied to Claims 1 and 11. These rejections are respectfully traversed.

As previously stated, Oh fails to teach each and every element recited by Claim 1. In addition, applicant respectfully submits that Oh fails to suggest each and every element recited by Claim 1. Itoh fails to overcome the significant shortcomings of Itoh to disclose or suggest the features of amended Claim 1.

Itoh is directed to an acoustic noise suppressor which suppresses signals other than speech signals or the like. (Itoh, column 1, lines 4-8.) In addition, Itoh teaches a

noise suppressor including an analysis/discrimination part **20**. (Itoh, column 4, lines 51-54.) Part **20** comprises an LPC analysis part **22**, an autocorrelation analysis part **23**, a maximum value detecting part **24** and a speech/non-speech identification part **25**. (Itoh, column 4, lines 55-58.) Further, Itoh teaches that part **30** includes a psychoacoustically weighted substration part **34** for multiplying a noise spectrum Sn(f) by a psychoacoustic weighting coefficient W(f) and subtracting the psychoacoustically weighted noise spectrum from spectrum S(f) provided from a frequency analysis part **31**. (Itoh, column 5, lines 8-13.) Nowhere does Itoh disclose or suggest a filter operating in a full band while a control circuit for adjusting the coefficient of the filter operating in a subband. Therefore, for these reasons, Claim 1 is believed to be patentably distinguished over the combination of Oh and Itoh because the references do not disclose or suggest the presently claimed subject matter.

Claims 3 and 6-10 depend from Claim 1. Therefore, Claims 3 and 6-10 include the features of Claim 1. Thus, the comments presented below relating to Claim 1 apply equally to Claims 3 and 6-10. For these reasons, Claims 3 and 6-10 are believed to be patentably distinguished over the combination of Oh and Itoh because the references do not disclose or suggest the presently claimed subject matter.

As previously stated, Oh fails to teach each and every element recited by Claim 11. In addition, applicant respectfully submits that Oh fails to suggest each and every element recited by Claim 11. Itoh fails to overcome the significant shortcomings of Itoh to disclose or suggest the features of amended Claim 11.

As previously stated, Itoh is directed to an acoustic noise suppressor which suppresses signals other than speech signals or the like. In addition, Itoh teaches a

noise suppressor including an analysis/discrimination part **20**. Nowhere does Itoh disclose or suggest a filtering is carried out in the full band while the determination of the coefficients is carried out in the subband. Therefore, for these reasons, Claim 11 is believed to be patentably distinguished over the combination of Oh and Itoh because the references do not disclose or suggest the presently claimed subject matter.

Claims 13, 14, and 21 depend from Claim 11. Therefore, Claims 13, 14, and 21 include the features of Claim 11. Thus, the comments presented below relating to Claim 11 apply equally to Claims 13, 14, and 21. For these reasons, Claims 13, 14, and 21 are believed to be patentably distinguished over the combination of Oh and Itoh because the references do not disclose or suggest the presently claimed subject matter.

Applicant respectfully submits that the teachings of Oh and Itoh, either alone or in combination, do not teach or suggest each and every feature of the present subject matter, and therefore that Claims 3, 6-10, 13, 14, and 21 are not obvious in view of the Oh and Itoh. Applicant, therefore, respectfully requests that the rejection of Claims 3, 6-10, 13, 14, and 21 under 35 U.S.C. § 103(a) be withdrawn and the claims allowed at this time.

CONCLUSION

In light of the above amendments and remarks, it is respectfully submitted that the present application is now in proper condition for allowance, and an early notice to such effect is earnestly solicited.

If any small matter should remain outstanding after the Patent Examiner has had an opportunity to review the above Remarks, the Patent Examiner is respectfully requested to telephone the undersigned patent attorney in order to resolve these matters and avoid the issuance of another Official Action.

DEPOSIT ACCOUNT

The Commissioner is hereby authorized to charge any fees associated with the filing of this correspondence to Deposit Account No. **50-0426**.

Respectfully submitted,

JENKINS, WILSON & TAYLOR, P.A.

Date: December 21, 2004

By: Richard E Jenkins

Richard E. Jenkins
Registration No. 28,428
Customer No: 25297

REJ/BJO/gwc

1406/14

ELSEVIER

Signal Processing 64 (1998) 21–32

SIGNAL PROCESSING

Combined acoustic echo control and noise reduction for hands-free telephony

S. Gustafsson, R. Martin, P. Vary*

Institute of Communication Systems and Data Processing, Aachen University of Technology, RWTH Aachen, Muffeterweg 3, D-52056 Aachen, Germany

Received 9 October 1997

Abstract

In this paper we propose an algorithm for combined acoustic echo control and noise reduction. The algorithm is developed on the basis of a minimum mean-square error criterion and consists of a (conventional) echo canceller and an additional noise and residual echo reduction filter. A special feature of the algorithm is the procedure for estimating the residual echo power spectral density which relies on the assumption that the phase of the estimated echo is approximately equal to the phase of the true echo. This assumption is verified by experimental results. The residual echo power density estimate and the noise power density estimate are then adaptively combined and used as an argument for some spectral weighting rule such that the residual echo is attenuated and effectively masked by a low level of intentionally left background noise. The paper concludes with experimental results for a typical car environment. © 1998 Elsevier Science B.V. All rights reserved.

Zusammenfassung

In diesem Artikel wird ein Algorithmus für die gemeinsame Reduktion von akustischen Echos und von Störgeräuschen vorgeschlagen. Der Algorithmus wird auf der Basis des Kriteriums kleinster mittlerer quadratischer Fehler entwickelt und besteht aus einem Echokompensator und einem zusätzlichen Störgeräusch- und Restechoreduktionsfilter. Eine spezielle Eigenschaft des Algorithmus besteht in der Schätzung des Leistungsdichtespektrums des Restechos, die auf der Annahme beruht, daß die Phase des geschätzten Echos ungefähr der Phase des wahren Echos entspricht. Diese Annahme wird experimentell bestätigt. Der Schätzwert für die Restecholeistungsdichte und die Störgeräuschleistungsdichte werden adaptiv kombiniert und als Argument für eine spektrale Gewichtungsregel verwendet, so daß das Restecho abgeschwächt wird und effektiv von dem verbleibenden Reststörgeräusch maskiert wird. Der Artikel schließt mit experimentellen Ergebnissen aus einer Kraftfahrzeugumgebung. © 1998 Elsevier Science B.V. All rights reserved.

Résumé

Nous proposons dans cet article un algorithme pour le contrôle de l'écho acoustique et la réduction du bruit simultanés. Cet algorithme est développé sur la base d'un critère d'erreur quadratique moyenne et consiste en un

---

* Corresponding author. Fax: +49 241 8888186; e-mail: vary@ind.rwth-aachen.de.

PII S0165-1684(97)00173-4

annulateur d'écho (conventionnel) et un filtre additionnel de réduction de bruit et d'écho résiduel. Un trait spécial de cet algorithme est la procédure d'estimation de la densité spectrale de puissance de l'écho résiduel qui repose sur l'hypothèse que la phase de l'écho estimé est approximativement égale à la phase du vrai écho. Cette hypothèse est confirmée par des résultats expérimentaux. L'estimée de la densité spectrale de puissance de l'écho résiduel et l'estimée de la densité spectrale de puissance du bruit sont alors combinées de manière adaptative et utilisées comme arguments d'un règle de pondération spectrale de telle sorte que l'écho résiduel est atténué et effectivement masqué par un niveau réduit de bruit de fond intentionnellement conservé. Cet article se conclue par des résultats expérimentaux pour un environnement automobile typique.

Keywords: Acoustic echo control; Noise reduction; Wiener filter; Psychoacoustics

1. Introduction

The problem of combined acoustic echo cancellation and noise reduction has found considerable interest recently. This interest is fueled by applications in mobile communications where both acoustic echo cancellation and noise reduction are necessary to achieve sufficient quality of the transmitted speech signal. The realization of such a combined system is, however, a challenging task. The difficulties of acoustic echo cancellation are mainly due to the high computational complexity of the echo canceller and influences which disturb the adaptation of the canceller such as ambient noise, near end speech, and variations of the acoustic environment. In mobile applications where all these factors play a significant role, it is difficult to reach the echo attenuation as required by ITU and ETSI recommendations, see e.g. [1,2], with an echo canceller alone. To achieve sufficient echo reduction, additional voice controlled attenuators or a nonlinear processing device, e.g. a center clipper, can be inserted into the signal paths which in turn limits the double talk capability of the hands-free system [3] or produces noticable nonlinear distortions.

The noise reduction task is also not easily solved since in the typical reverberant environment no 'noise only' reference signal can be obtained which is sufficiently correlated to the noise within the microphone signal. Besides this principal restriction, most automobile manufacturers and mobile communication equipment suppliers favour single microphone solutions. Thus, although a multi-microphone system might yield better noise reduction, a single microphone spectral weighting ('spectral subtraction') technique is often prefered. These methods, however, have well-known disadvantages such as limited performance at low SNR values and artificial sounding residual noise.

In this paper we will present an algorithm for combined acoustic echo and noise reduction and summarize some of our research results. The algorithm is developed on the basis of a minimum mean-square error criterion. We show that acoustic echo control and noise reduction can be combined in a true synergy and that the combined approach will ease at least some of the above problems. The algorithm utilizes a conventional echo canceller of reduced order and an adaptive noise and residual echo reduction filter. We do not strive to achieve complete cancellation with the echo canceller alone but we rather use an echo canceller of reduced order to decrease the computational complexity and improve the overall robustness of the adaptation process. The required echo reduction is then achieved by further attenuating the residual echo with a combined noise and residual echo reduction filter in the sending path of the hands-free telephone. Similar to the well-known spectral subtraction noise reduction technique [4,5] our approach requires an estimate of the power spectral densities of the ambient noise and the residual echo after echo compensation. The key issue of this paper is to show how an estimate of the power spectral density of the residual echo might be obtained.

The echo cancellation and noise reduction problem has been addressed independently for many years (see e.g. [6–10] for reviews of these methods). In the last years it has been recognized, however, that the echo control and noise reduction problem can be tackled in a combined approach [11–19]. It has been shown that the combined

treatment yields algorithms which deliver better performance at less computational costs than systems based on separate algorithms [14,17,18]. Previous methods, however, do not feature the explicit residual echo estimation.

The remainder of this paper is organized as follows. In the next section we develop the basis of our approach, i.e. the minimum mean-square solution to the combined problem. Section 3 discusses our algorithm in detail with special emphasis on the estimation of the residual echo power spectral density and the computation of the spectral weighting function. Finally, Section 4 presents and discusses our experimental results.

2. An optimal solution to the combined problem

Fig. 1 depicts the basic scenario for hands-free telephony. We assume that all signals are band-limited, digitized and that the microphone signal $y(k)$ is a linear combination of the near end speech $s(k)$, the near end ambient noise $n(k)$, and the echo signal $d(k)$. The echo signal $d(k)$ typically consists of a component which is linearly related to the loudspeaker signal $x(k)$ and a component which is the result of nonlinear distortions of the loudspeaker signal. In a stationary scenario the former can be, at least theoretically, identified and compensated by an echo canceller. In a practical implementation where the order of the canceller, the time variance of the acoustic environment, and the ambient noise have a significant influence, the linearly related

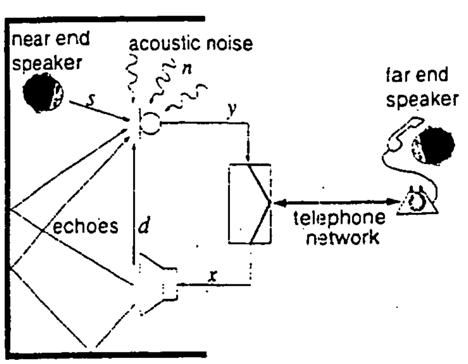

Fig. 1. Basic hands-free telephony scenario.

component cannot be completely eliminated by the canceller and will thus, together with the nonlinear component, contribute to the residual echo.

To combine acoustic echo cancellation with residual echo and noise reduction it must be asked in which order these processing operations should be performed. Although there are good arguments in favour of processing first the noise reduction, our considerations and experimental results clearly show that the configuration where the echo compensation (EC) preceeds the echo and noise reduction (ENR) is preferable. The main advantage of the EC/ENR configuration is that the noise reduction has not to cope with the disturbing echo signal as it is present in the microphone signal and that there is no time varying noise reduction filter in the echo path. Besides that, if the echo canceller does not deliver sufficient echo attenuation, the residual echo can be treated similar to the background noise signal and can be further attenuated by the noise reduction filter. This idea is successfully exploited in a frequency selective echo reduction technique, called 'echo shaping' [20,21], which does not require complete cancellation of the echo by the echo canceller and is easily combined with a noise reduction filter. A disadvantage of the EC/ENR configuration is that the echo canceller has to process noisy signals. As a result, algorithms have been proposed where besides the noise reduction filter in the sending path, noise reduced signals are used to adapt the echo canceller [13,17].

For the derivation of the optimal solution in the minimum mean-square error sense, we assume that all signals are stationary and that the estimated signal $\hat{s}(k)$ at the output of the combined system is the result of linearly filtering the far end signal $x(k)$ and the microphone signal $y(k)$, i.e.

$$\hat{s}(k) = y(k)*w_1(k) + x(k)*w_2(k), \tag{1}$$

where $w_1(k)$ and $w_2(k)$ are the impulse responses of two unconstrained (IIR) or constrained (FIR) adaptive filters and $*$ denotes the convolution operation.

It is interesting to note that the very general approach of minimizing the mean-square error $\mathcal{E}\{(s(k) - \hat{s}(k))^2\}$ ($\mathcal{E}\{\cdot\}$ denotes the expectation operator) suggests an algorithm for the combined system which first computes an echo compensated

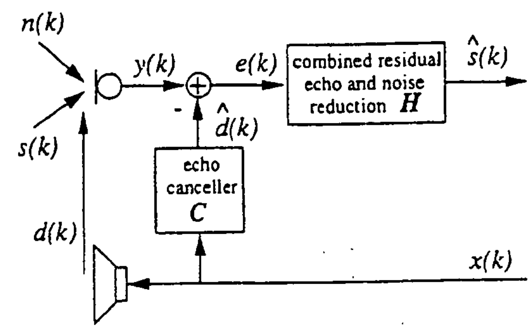

Fig. 2. Block diagram of the acoustic echo cancellation and noise reduction system.

signal $e(k)$ and then performs a combined noise and residual echo reduction [18,22].

In fact, the unconstrained (IIR) optimal solution is given by (see Appendix A)

$$\hat{s}(k) = \left(y(k) - x(k) * \mathcal{F}^{-1}\left\{\frac{R_{xy}(\Omega)}{R_{xx}(\Omega)}\right\}\right) * \mathcal{F}^{-1}\left\{\frac{R_{ss}(\Omega)}{R_{yy}(\Omega) - R_{yx}(\Omega)R_{xx}^{-1}(\Omega)R_{xy}(\Omega)}\right\}, \quad (2)$$

where $\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse Fourier transform of discrete time signals and $R_{xx}(\Omega)$, $R_{xy}(\Omega)$, $R_{yx}(\Omega)$, $R_{yy}(\Omega)$, $R_{ss}(\Omega)$ denote the (cross-) power spectral densities of the signals in the subscripts. According to the above solution, the combined problem is fully separable into the optimal echo canceller with a frequency response $C(\Omega) = R_{xy}(\Omega)R_{xx}^{-1}(\Omega)$ and a combined residual echo and noise reduction filter with frequency response $H(\Omega) = W_1(\Omega) = R_{ss}(\Omega)[R_{yy}(\Omega) - R_{yx}(\Omega)R_{xx}^{-1}(\Omega)R_{xy}(\Omega)]^{-1}$. A block diagram of the resulting system is shown in Fig. 2. If the nonlinear distortions of the loudspeaker signal are neglected, the echo signal $d(k)$ is linearly related to the loudspeaker signal $x(k)$ and the unconstrained optimal echo canceller will deliver a perfect estimate of the echo signal. In this case we find that $R_{yy}(\Omega) - R_{yx}(\Omega)R_{xx}^{-1}(\Omega)R_{xy}(\Omega) = R_{ss}(\Omega) + R_{nn}(\Omega)$ is valid and the echo and noise reduction postfilter reduces to the well-known Wiener filter for a signal with additive noise. In general, and especially in the constrained FIR case, however, a residual echo component $b(k) = d(k) - \hat{d}(k)$ will remain after the non-perfect echo compensation.

In case that the echo signal $d(k)$ cannot be completely cancelled (because of a non-perfect canceller and/or additional non-linear components), the residual echo and noise reduction filter $H(\Omega)$ is given by

$$H(\Omega) = \frac{R_{ss}(\Omega)}{R_{ss}(\Omega) + R_{nn}(\Omega) + R_{bb}(\Omega)}, \quad (3)$$

where $R_{bb}(\Omega)$ denotes the power spectral density of the residual echo $b(k)$. Similar to well-known noise reduction techniques, the estimation of the optimal filter $H(\Omega)$ requires estimates of the noise power spectral density $R_{nn}(\Omega)$ and the residual echo power spectral density $R_{bb}(\Omega)$. It should be noted that in contrast to the ambient noise $n(k)$, the residual echo $b(k)$ is a speech-like signal. Thus, the estimation procedures for $R_{nn}(\Omega)$ and $R_{bb}(\Omega)$ are entirely different. Furthermore, unlike the optimal unconstrained solution for the above stationary scenario, any real adaptive implementation using a limited amount of data will be sensitive to ambient noise. Since there is no noise reduction before echo cancellation, the optimal solution as suggested by Eq. (2) requires a very robust echo canceller, especially in a car environment. If the residual echo power spectral density is known with sufficient accuracy, performance deficiencies of the canceller can be counterbalanced by the filter $H(\Omega)$.

3. An algorithm for combined echo and noise reduction

In this section we present an algorithm which closely follows the minimum mean square approach of the previous section. Since the main focus of this section is on the adaptation of the optimal filter $H(\Omega)$, we will not discuss the echo canceller in detail. The echo canceller in our combined system was developed by Antweiler [23] and utilizes an adaptive step-size control algorithm due to Frenzel [24]. This canceller has been proven to be very robust even in noisy and time variant environments. The canceller is described in detail in [23].

3.1. SNR estimation and weighting rules

State-of-the-art noise reduction algorithms are based on a priori and a posteriori signal-to-noise ratio (SNR) estimates [5,25]. To facilitate the subsequent discussion of estimation procedures, we introduce a frame index (m) and discrete frequencies $\Omega_i = 2\pi i/M$, $i \in \{0, 1, 2, 3, \ldots, M-1\}$ with M being the FFT frame size. $R_{ss}^{(m)}(\Omega_i)$, $R_{nn}^{(m)}(\Omega_i)$ and $R_{bb}^{(m)}(\Omega_i)$ denote the power spectral densities – or short-term estimates thereof – of the signals in the subscripts for the mth frame. The a priori SNR for the combined residual echo and noise reduction problem is then given by

$$\mathrm{SNR}_{b+n}^{s,(m)}(\Omega_i) = \frac{R_{ss}^{(m)}(\Omega_i)}{R_{nn}^{(m)}(\Omega_i) + R_{bb}^{(m)}(\Omega_i)}, \tag{4}$$

which can be rewritten in terms of the individual SNR values related to ambient noise $n(k)$ and the residual echo $b(k)$,

$$\mathrm{SNR}_{b+n}^{s,(m)}(\Omega_i) = \frac{1}{[\mathrm{SNR}_n^{s,(m)}(\Omega_i)]^{-1} + [\mathrm{SNR}_b^{s,(m)}(\Omega_i)]^{-1}} \tag{5}$$

where $\mathrm{SNR}_n^{s,(m)}(\Omega_i)$ and $\mathrm{SNR}_b^{s,(m)}(\Omega_i)$ are given by

$$\mathrm{SNR}_n^{s,(m)}(\Omega_i) = \frac{R_{ss}^{(m)}(\Omega_i)}{R_{nn}^{(m)}(\Omega_i)} \tag{6}$$

and

$$\mathrm{SNR}_b^{s,(m)}(\Omega_i) = \frac{R_{ss}^{(m)}(\Omega_i)}{R_{bb}^{(m)}(\Omega_i)}. \tag{7}$$

The optimal filter $H(\Omega)$ in Eq. (3) is now easily expressed in terms of the a priori SNR.

$$H^{(m)}(\Omega_i) = \frac{\mathrm{SNR}_{b+n}^{s,(m)}(\Omega_i)}{\mathrm{SNR}_{b+n}^{s,(m)}(\Omega_i) + 1}. \tag{8}$$

Similar to [5,18,19] the individual a priori SNR values are estimated by a 'decision directed' approach. The a priori SNR related to the ambient noise $n(k)$ is given by

$$\mathrm{SNR}_n^{s,(m)}(\Omega_i) = (1-\alpha_n)P(\mathrm{SNR}_n^{e,(m)}(\Omega_i) - 1) + \alpha_n \frac{|H^{(m-1)}(\Omega_i)E^{(m-1)}(\Omega_i)|^2}{R_{nn}^{(m)}(\Omega_i)}, \tag{9}$$

where $P(x) = \frac{1}{2}(|x| + x)$. $\mathrm{SNR}_n^{e,(m)}(\Omega_i)$ denotes the a posteriori SNR with respect to the ambient noise $n(k)$,

$$\mathrm{SNR}_n^{e,(m)}(\Omega_i) = \frac{|E^{(m)}(\Omega_i)|^2}{R_{nn}^{(m)}(\Omega_i)}, \tag{10}$$

α_n is a step-size parameter, and $E^{(m)}(\Omega_i)$ the discrete Fourier transform of the compensated signal $e(k)$. A similar expression holds for the SNR related to the residual echo $b(k)$.

3.2. Power spectral density estimation

For the combined reduction of residual echo and noise, separate estimations of the power spectral densities of the background noise $n(k)$ and the residual echo $b(k)$ have to be performed, as the characteristics of the speech-like residual echo differs very much from that of the noise. The noise power spectral density $R_{nn}^{(m)}(\Omega_i)$ can be estimated by the 'minimum statistics' or 'spectral minima tracking' methods as outlined in [26,27]. These methods have the advantage that the noise power spectral density is estimated continuously, eliminating the need for a voice activity detector. They are also able to track slow variations of the noise power density, which is vital for the noise reduction algorithm to perform well if $R_{nn}^{(m)}(\Omega_i)$ is changing during speech activity.

The estimation of the residual echo power spectral density $R_{bb}^{(m)}(\Omega_i)$ is much more involved since $b(k)$ is a speech-like signal and thus stationary over time periods of only 20–500 ms. The main idea of our estimation procedure is therefore to derive $R_{bb}^{(m)}(\Omega_i)$ from quantities which change less rapidly over time. We achieve this by modelling the residual echo as the output of a linear system with the echo $d(k)$ as its input and by assuming that the transfer function $F^{(m)}(\Omega_i)$ of this possibly non-causal and time varying system is statistically independent from the input signal $d(k)$. This assumption is, of course, not entirely fulfilled, because the echo cancellation, which is modelled by $F^{(m)}(\Omega_i)$, is controlled by an adaptive algorithm which itself is dependent on the echo $d(k)$. But over short time intervals, when the room impulse response does not change

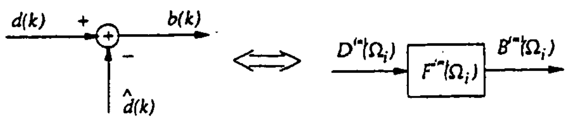

Fig. 3. Interpretation of the echo compensation as a transfer function $F^{(m)}(\Omega_i)$.

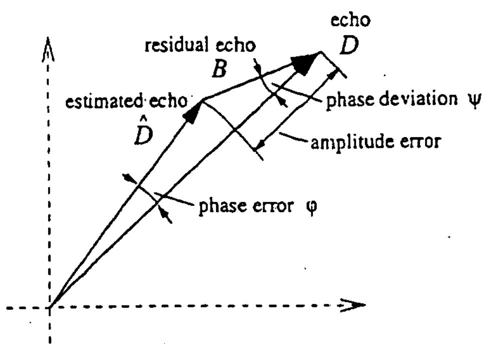

Fig. 4. Frequency domain vector diagram of the echo $D^{(m)}(\Omega_i)$, the estimated echo $\hat{D}^{(m)}(\Omega_i)$ and the residual echo $B^{(m)}(\Omega_i)$.

rapidly and the echo canceller works in a steady-state condition, one can assume $F^{(m)}(\Omega_i)$ to be almost constant and thus to be independent from $d(k)$.

The approach is illustrated in Fig. 3. In terms of short-term frame oriented spectral analysis it leads to the identities

$$B^{(m)}(\Omega_i) = D^{(m)}(\Omega_i) - \hat{D}^{(m)}(\Omega_i), \tag{11}$$

$$B^{(m)}(\Omega_i) = F^{(m)}(\Omega_i) D^{(m)}(\Omega_i). \tag{12}$$

For our estimation procedure, we also assume (and verify by measurement) that the transfer function $F^{(m)}(\Omega_i)$ can be approximated by a real valued function. In Fig. 4 the vectors $D^{(m)}(\Omega_i)$, $\hat{D}^{(m)}(\Omega_i)$ and $B^{(m)}(\Omega_i)$ are plotted for a given frequency Ω_i in the complex plane. The misalignment between $D^{(m)}(\Omega_i)$ and $\hat{D}^{(m)}(\Omega_i)$ can be expressed in the magnitude error $|D^{(m)}(\Omega_i)| - |\hat{D}^{(m)}(\Omega_i)|$ and the phase error $\varphi = \arg\{D^{(m)}(\Omega_i)\} - \arg\{\hat{D}^{(m)}(\Omega_i)\}$. To justify the assumption, we note that – for some fixed phase error φ – the more the magnitude of the estimated echo $|\hat{D}^{(m)}(\Omega_i)|$ deviates from the magnitude of the true echo $|D^{(m)}(\Omega_i)|$, the smaller the phase deviation ψ of vectors $D^{(m)}(\Omega_i)$ and $B^{(m)}(\Omega_i)$ is. Thus, in this case our

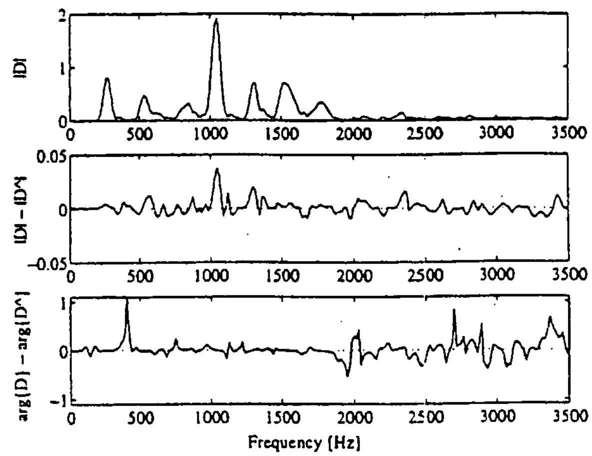

Fig. 5. Time domain echo compensation: the magnitude of the echo, the magnitude error, and the phase error (in radians), for a sample speech frame.

assumption will give a good approximation, which will be even better the smaller φ is. It has been verified by simulations that due to the compensation mechanism the phase error φ is indeed small at frequencies where the echo is present. A sample frame of $|D^{(m)}(\Omega_i)|$, of the magnitude error, and of the phase error are shown in Fig. 5. A large phase error can only be found at frequencies where $|D^{(m)}(\Omega_i)|$ is very small.

If the magnitude error between $D^{(m)}(\Omega_i)$ and $\hat{D}^{(m)}(\Omega_i)$ is very small, the phase of $B^{(m)}(\Omega_i)$ might be entirely different from the phase of $D^{(m)}(\Omega_i)$. This, however, does not pose a serious problem since the residual echo is then small and we are only interested in the magnitude of the residual echo. The magnitude of $B^{(m)}(\Omega_i)$ will then be underestimated, which results in less echo reduction but not in additional distortions of the near end speech signal.

From Eq. (12) we can write the power spectral density of the residual echo, $R_{bb}^{(m)}(\Omega_i)$, as a function of the power spectral density of the echo, $R_{dd}^{(m)}(\Omega_i)$,

$$\begin{aligned} R_{bb}^{(m)}(\Omega_i) &= |F^{(m)}(\Omega_i)|^2 R_{dd}^{(m)}(\Omega_i) \\ &= (F^{(m)}(\Omega_i))^2 R_{dd}^{(m)}(\Omega_i). \end{aligned} \tag{13}$$

By combining Eqs. (11) and (12), both $R_{dd}^{(m)}(\Omega_i)$ and $R_{bb}^{(m)}(\Omega_i)$ can be written as functions of the transfer function $F^{(m)}(\Omega_i)$ and the power spectral density of

the estimated echo, $R_{\hat{d}\hat{d}}^{(m)}(\Omega_i)$.

$$R_{dd}^{(m)}(\Omega_i) = \frac{1}{(1 - F^{(m)}(\Omega_i))^2} R_{\hat{d}\hat{d}}^{(m)}(\Omega_i), \tag{14}$$

$$R_{bb}^{(m)}(\Omega_i) = \left(\frac{F^{(m)}(\Omega_i)}{1 - F^{(m)}(\Omega_i)}\right)^2 R_{\hat{d}\hat{d}}^{(m)}(\Omega_i), \tag{15}$$

which are well defined for any $F^{(m)}(\Omega_i) \neq 1$, which practically means that the echo canceller delivers a non-zero echo estimate. The problem of estimating $R_{bb}^{(m)}(\Omega_i)$ is then converted into the estimation of the transfer function $F^{(m)}(\Omega_i)$.

If neither near end speech nor near end noise is present, i.e. a noise-free single talk situation where $y(k) = d(k)$ and $e(k) = b(k)$, $F^{(m)}(\Omega_i)$ can be calculated from Eq. (12),

$$(F^{(m)}(\Omega_i))^2 = \frac{R_{bb}^{(m)}(\Omega_i)}{R_{\hat{d}\hat{d}}^{(m)}(\Omega_i)}. \tag{16}$$

However, as this situation seldom prevails, another solution must be found.

Assuming statistical independence between the near end speech $s(k)$, the noise $n(k)$, and the echo $d(k)$ or the residual echo $b(k)$, we can write the power spectral densities of the microphone signal $y(k)$ and the compensated signal $e(k)$ as

$$R_{yy}^{(m)}(\Omega_i) = R_{ss}^{(m)}(\Omega_i) + R_{nn}^{(m)}(\Omega_i) + R_{dd}^{(m)}(\Omega_i), \tag{17}$$

$$R_{ee}^{(m)}(\Omega_i) = R_{ss}^{(m)}(\Omega_i) + R_{nn}^{(m)}(\Omega_i) + R_{bb}^{(m)}(\Omega_i). \tag{18}$$

Combining the above equations with Eqs. (14) and (15), we arrive at an expression for estimating $F^{(m)}(\Omega_i)$, which can now be calculated from measurable quantities (see Appendix B).

$$F^{(m)}(\Omega_i) = \frac{R_{yy}^{(m)}(\Omega_i) - R_{ee}^{(m)}(\Omega_i) - R_{\hat{d}\hat{d}}^{(m)}(\Omega_i)}{R_{yy}^{(m)}(\Omega_i) - R_{ee}^{(m)}(\Omega_i) + R_{\hat{d}\hat{d}}^{(m)}(\Omega_i)}. \tag{19}$$

Eq. (19) is only valid if $R_{\hat{d}\hat{d}}^{(m)}(\Omega_i) \neq 0$. Under some circumstances, for example, when the estimated echo power spectral density is very weak compared to the power spectral density of the microphone signal, Eq. (19) can, owing to estimation errors and finite numerical accuracy, lead to wrong results. Therefore, potential errors must be excluded from the calculation. In our algorithm this is achieved in four steps:

1. Limit $F^{(m)}(\Omega_i)$ to some reasonable range $[F_{min}, F_{max}]$.
2. Split the frequency range in $N < M$ subbands.
3. In each subband, calculate the mean value $\bar{F}_n^{(m)}$ of those $F_n^{(m)}(\Omega_i)$ where $R_{\hat{d}\hat{d}}^{(m)}(\Omega_i)$ is not too small.
4. At each frequency Ω_i, set $\tilde{F}^{(m)}(\Omega_i)$ to the corresponding mean value $\bar{F}_n^{(m)}$.

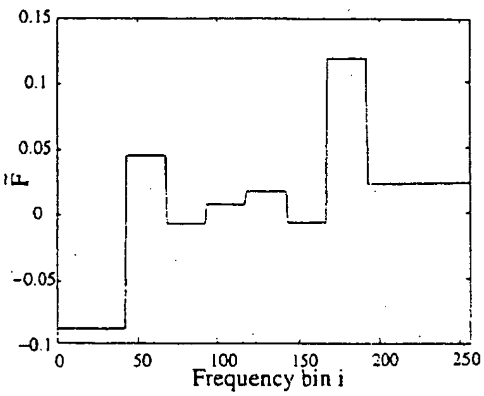

Fig. 6. A sample of the estimated transfer function $\tilde{F}^{(m)}(\Omega_i)$.

The transfer function $\tilde{F}^{(m)}(\Omega_i)$ estimated this way will then be used for the estimation of $R_{bb}(\Omega_i)$ using Eq. (15). It will possess a frequency-dependent step-shape as illustrated in Fig. 6.

3.3. Limiting of estimated SNR

With the residual echo power spectral density $R_{bb}^{(m)}(\Omega_i)$ and the noise power spectral density $R_{nn}^{(m)}(\Omega_i)$ estimated as described in the previous section, we can now determine the SNR values and compute the spectral weighting coefficients. An improvement of the auditive impression can be accomplished if the SNR estimates of the residual echo and the noise are limited and balanced with respect to each other. Also, it is often desirable to leave a low level of natural sounding residual noise in the processed signal. This can be achieved by limiting the estimated a priori SNR to a minimum threshold T_n,

$$\begin{aligned} \widetilde{\mathrm{SNR}}_n^{s,(m)}(\Omega_i) &= \max(\mathrm{SNR}_n^{s,(m)}(\Omega_i), T_n), \\ \widetilde{\mathrm{SNR}}_n^{r,(m)}(\Omega_i) &= \max(\mathrm{SNR}_n^{r,(m)}(\Omega_i), T_n). \end{aligned} \tag{20}$$

Especially at SNR values below T_n the limiting will have a significant effect. Thus, the stronger the background noise level is, the higher the noise level in the processed signal will be. This is in fact an advantage, as speech enhancement in general performs less well in a low SNR-environment. The limiting prevents too high an attenuation and therefore it also reduces the distortions of the near end speech, which otherwise might lose intelligibility. A proper range is $T_n = 0.01$–0.1, where the chosen value eventually depends on criteria such as desired noise reduction and admissible speech distortion. If the threshold T_n is chosen too high, the amount of noise reduction will be very low; if it is chosen too low, the limiting will have almost no influence.

Now consider an equivalent limiting of the SNR refering to the residual echo using a constant threshold T_b. This will have the effect that some residual echo will always be left in the signal $\hat{s}(k)$. Of course, this is not desirable in a noise-free situation, as the echo might then be audible. However, when noise is present, some limiting is necessary as otherwise the attenuation by the filter H might be too high, leading to disturbing modulations of the residual noise whenever the far end speaker is active.

We therefore propose to attenuate the residual echo $b(k)$ where it is most likely not masked by the residual noise. In the processed signal $\hat{s}(k)$ only the near end speech and an attenuated, natural sounding background noise should be audible, but no echo. This can be achieved by a frequency dependent limiting with a threshold $T_b^{(m)}(\Omega_i)$,

$$\widetilde{\mathrm{SNR}}_b^{a,(m)}(\Omega_i) = \max(\mathrm{SNR}_b^{a,(m)}(\Omega_i), T_b^{(m)}(\Omega_i)),$$
$$\widetilde{\mathrm{SNR}}_b^{e,(m)}(\Omega_i) = \max(\mathrm{SNR}_b^{e,(m)}(\Omega_i), T_b^{(m)}(\Omega_i)), \tag{21}$$

where $T_b^{(m)}(\Omega_i)$ is a function of the chosen threshold T_n and of the power spectral densities of the residual echo and the noise, for example

$$T_b^{(m)}(\Omega_i) = \frac{2T_n}{1 + R_{bb}^{(m)}(\Omega_i)/R_{nn}^{(m)}(\Omega_i)}. \tag{22}$$

With this limiting function, $T_b^{(m)}(\Omega_i) \to 0$ if $R_{nn}^{(m)}(\Omega_i) \to 0$, thus permitting complete attenuation of the residual echo. When there is a strong noise present, which already masks the residual echo, $R_{nn}^{(m)}(\Omega_i) \gg R_{bb}^{(m)}(\Omega_i)$ leads to $T_b^{(m)}(\Omega_i) \approx 2T_n$, effectively preventing too high an attenuation. Finally, if $R_{nn}^{(m)}(\Omega_i) = R_{bb}^{(m)}(\Omega_i)$ then $T_b^{(m)}(\Omega_i) = T_n$ and the SNRs are all limited to the same level. In an idealized stationary condition this would lead to a combined SNR of exactly half the value of the individual SNRs (see Eq. (5)).

With the above adaptive limiting and the subsequent combination of the different a priori and a posteriori SNRs, the speech enhancement algorithm will work well over a wide range of input signal-to-noise conditions, effectively reducing the background noise and the residual echo with only minor impacts on the near end speech quality.

4. Experimental results

Our algorithm was evaluated in a car environment with single talk, double talk, and various ambient noise levels. The sample frequency was 8000 Hz. In the car, which had a reverberation time of about 70 ms, the combined system with an echo canceller of only $N_c = 200$ filter taps gives satisfactory performance. The residual echo and noise reduction filter $H(\Omega_i)$ was realized in the frequency domain by means of a framewise processing with a 512 point FFT with 50% overlap. The frames consisted of 256 data samples multiplied by a Hamming window and were zero padded to the full FFT length.

All experimental conditions were evaluated using informal listening tests and the instrumental assessment method as described below.

4.1. Instrumental assessment

Our evaluation method is based on a separate processing of the acoustic echo and the near end signal [28,29]. This evaluation scheme requires, however, that the near end speech and noise signals are recorded independently of the echo signal. The simulation setup is shown in Fig. 7.

Based on the signals shown in Fig. 7 the following measures can be defined:

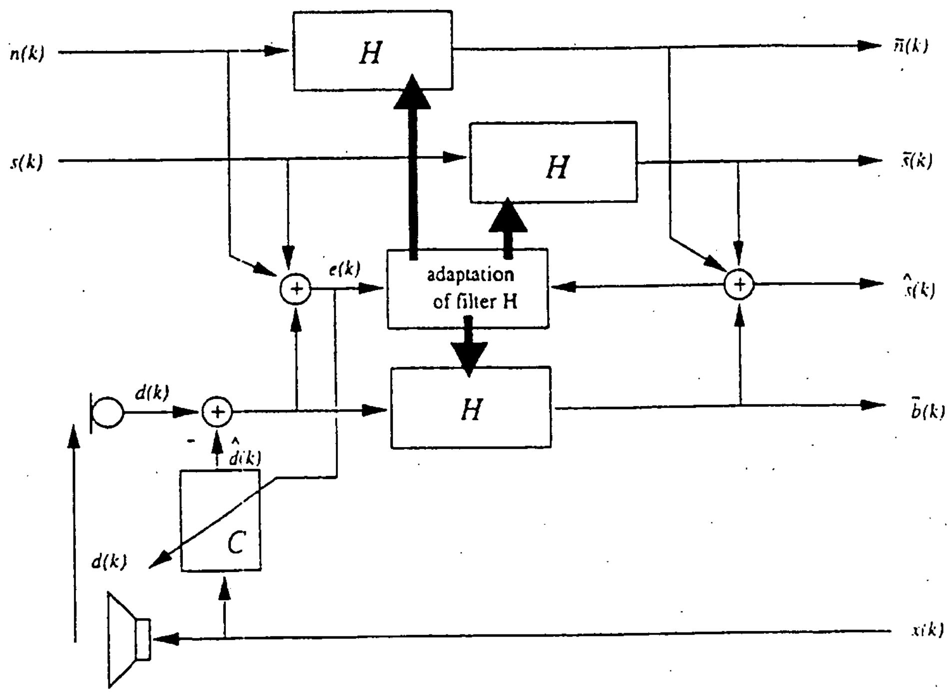

Fig. 7. Signal model for the instrumental evaluation of the combined algorithm.

- the time average of the echo return loss enhancement of the compensator C,

$$\mathrm{ERLE}_C = \frac{10}{k_{max} - k_1} \sum_{k=k_1+1}^{k_{max}} \log_{10}\left(\frac{\mathcal{E}\{d^2(k)\}}{\mathcal{E}\{b^2(k)\}}\right), \tag{23}$$

where $k_1 + 1$ is the index of the first sample and k_{max} is the index of the last sample of the measurement;

- the time average of the echo attenuation of the combined system (compensator C + filter H),

$$\mathrm{ERLE}_{CH} = \frac{10}{k_{max} - k_1} \sum_{k=k_1+1}^{k_{max}} \log_{10}\left(\frac{\mathcal{E}\{d^2(k)\}}{\mathcal{E}\{\tilde{b}^2(k)\}}\right), \tag{24}$$

where $\tilde{b}(k)$ is the residual echo $b(k)$ filtered with the filter H.

- the distortion of the near end signal caused by the filter H as measured by the segmental SNR,

$$\mathrm{SEGSNR} = \frac{1}{K_{\mathrm{SNR}>0}} \sum_{m=0}^{K-1} \max(\mathrm{SNR}_{seg}(m), 0), \tag{25}$$

with

$$\mathrm{SNR}_{seg}(m) := 10 \log_{10}\left(\frac{\sum_{i=mN}^{mN+N-1} s^2(i)}{\sum_{i=mN}^{mN+N-1} (\tilde{s}(i) - s(i - N_H))^2}\right), \tag{26}$$

where N_H is the delay caused by the filter H. N is the segment length. K is the total number of segments, and $K_{\mathrm{SNR}>0}$ is the number of frames with $\mathrm{SNR}_{seg} > 0$;

- the noise reduction NR of the combined residual echo and noise reduction filter H,

$$\mathrm{NR} = \frac{10}{k_{max} - k_1} \sum_{k=k_1+1}^{k_{max}} \log_{10}\left(\frac{\mathcal{E}\{\tilde{n}^2(k)\}}{\mathcal{E}\{n^2(k)\}}\right). \tag{27}$$

The expected values $\mathcal{E}\{\cdot\}$ are computed as ensemble averages across 8–16 phonetically balanced sentences. Since $H(\Omega)$ is a linear phase filter the SEGSNR criterion measures only the amplitude distortions of the near end signal.

4.2. Single talk experiments

The single talk situation was evaluated using 16 phonetically balanced sentences and additional car noise. Three different echo-to-noise ratios were considered, $SNR_n^d = 0$, 10 and 25 dB. In Fig. 8, the mean echo return loss enhancement (ERLE) for the compensator ($ERLE_C$), for the combined system ($ERLE_{CH}$), and the noise reduction is plotted as a function of the echo-to-noise ratio at the microphone input. As expected, less noise in the microphone input results in a better echo attenuation. In the noiseless case, the overall echo attenuation is about 50 dB which is sufficient to fulfill ITU and ETSI recommendations. As the level of noise increases, less echo attenuation is necessary and desirable, since some of the residual echo is masked by the noise and too high an attenuation would result in a fluctuating residual noise signal with a negative impact on the perceived quality. Interestingly, the echo canceller still performs well at $SNR_n^d = 0$ dB. This is because of its relatively short length, which makes it more robust in noisy environments. The good performance is also a condition for the residual echo reduction to work satisfactorily.

4.3. Double talk experiments

In our double talk experiments eight phonetically balanced sentences were used both for the

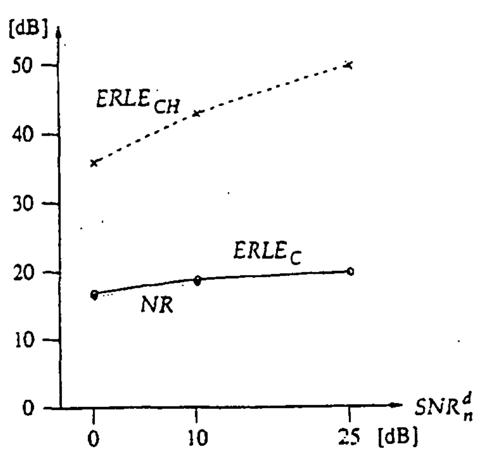

Fig. 8. Simulation results for single talk situations. $ERLE_C$: ERLE with compensator C; $ERLE_{CH}$: ERLE with compensator C and filter H; NR: noise reduction.

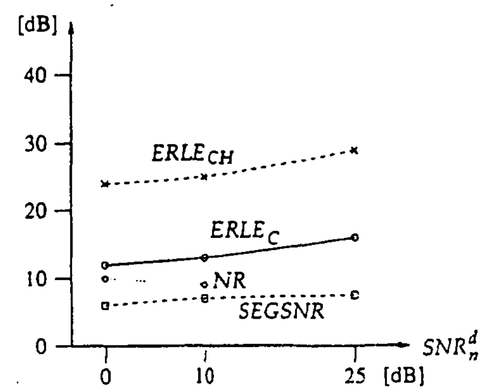

Fig. 9. Simulation results for double talk situations. $ERLE_C$: ERLE with compensator C; $ERLE_{CH}$: ERLE with compensator C and filter H; NR: noise reduction; SEGSNR: segmental SNR for near end speech.

near end speech and the far end speech. Fig. 9 shows the ERLE of the compensator ($ERLE_C$), the ERLE of the combined system ($ERLE_{CH}$), the noise reduction NR, and the segmental SNR of the processed near end speech (SEGSNR) as a function of the echo-to-noise ratio SNR_n^d. The near end speech was of about the same power as the far end speech. Compared to the single talk case the overall echo reduction is now about 20 dB lower. Because of the double talk the compensator now converges much slower. Again, since the near end signals will mask some of the residual echo, a higher overall echo attenuation is not desirable, since it would only lead to more distortions of the near end speech signal.

Informal listening experiments confirm that the remaining echo is indeed almost unhearable. Of course, double talk results in light distortions of the near end speech signal, but since the far end speaker is active at the same time these will have only a minor effect on the perceived quality.

5. Conclusions

In this paper we have shown by theory how the tasks of acoustic echo attenuation and noise reduction can be combined. We propose a structure consisting of an acoustic echo canceller which is followed by an adaptive postfilter. The postfilter is

able to attenuate not only the background noise, but also the residual echo left by the echo canceller. Thus, the length of the echo canceller can be reduced.

In order to be applicable in the considered hands-free telephony environment, we demand from the system to perform a significant acoustic echo and noise reduction for a wide range of signal-to-noise conditions. Furthermore, a high near end speech quality and a natural sounding residual background noise are of great importance.

To achieve this the power spectral densities of the residual echo and the background noise, which have inherently different characteristics, are estimated separately. A further new feature of our algorithm is an adaptive combination of the separate estimates, such that a low level of background noise will remain. It has been found that by carefully balancing the residual echo and noise attenuation the psychoacoustic masking effects of the ear contribute to a significant improvement of the perceived quality of the processed near end signal.

The complexity of the algorithm depends mainly on the order N_C of the time domain echo canceller C. For $N_C = 200$, the echo canceller and the postfilter H will need approximately the same number of operations per input sample. The total number of operations is then estimated to be less than 20 MIPS using a typical DSP.

The extra delay introduced by the filter H depends on the implementation of the analysis/synthesis structure, especially the FFT-length and the zero padding, and is in our system about 256 samples.

Acknowledgements

The authors gratefully acknowledge the funding of this project by Deutsche Telekom AG. They thank Dr. H. Schütze, Dr. R. Zelinski, and E. Diedrich for their support. They also thank P. Jax who wrote some of the C++ code for the simulation software.

Appendix A. Derivation of the optimal structure for a combined noise and residual echo reduction filter

Differentiation the expectation $\mathcal{E}\{(s(k) - \hat{s}(k))^2\}$ with respect to the unknown filter coefficients, where

$$\hat{s}(k) = y(k)*w_1(k) + x(k)*w_2(k), \tag{A.1}$$

leads to the normal equations

$$\begin{aligned}\frac{\partial \mathcal{E}\{e^2(k)\}}{\partial w_1(i)} &= -r_{ys}(i) + w_1(i)*r_{yy}(i) + w_2(i)*r_{yx}(i)\\ &= 0,\\ \frac{\partial \mathcal{E}\{e^2(k)\}}{\partial w_2(i)} &= -r_{xs}(i) + w_1(i)*r_{xy}(i) + w_2(i)*r_{xx}(i)\\ &= 0 \quad \forall i,\end{aligned} \tag{A.2}$$

where $r_{xy}(i) = \mathcal{E}\{x(k)y(k+i)\}$ denotes the correlation function of the signals in the subscripts. Eq. (A.2) yields in the frequency domain

$$\begin{aligned}W_1(\Omega)R_{yy}(\Omega) + W_2(\Omega)R_{yx}(\Omega) &= R_{ys}(\Omega),\\ W_1(\Omega)R_{xy}(\Omega) + W_2(\Omega)R_{xx}(\Omega) &= R_{xs}(\Omega) = 0,\end{aligned} \tag{A.3}$$

which readily gives the desired result

$$\begin{aligned}W_1(\Omega) &= \frac{R_{ss}(\Omega)}{R_{yy}(\Omega) - R_{xx}^{-1}(\Omega)R_{xy}(\Omega)R_{yx}(\Omega)} = H(\Omega),\\ W_2(\Omega) &= -\frac{R_{xy}(\Omega)}{R_{xx}(\Omega)}W_1(\Omega).\end{aligned} \tag{A.4}$$

Appendix B. Derivation of Eq. (19)

Recall Eqs. (17) and (18)

$$R_{yy}^{(m)}(\Omega_i) = R_{ss}^{(m)}(\Omega_i) + R_{nn}^{(m)}(\Omega_i) + R_{dd}^{(m)}(\Omega_i),$$

$$R_{ee}^{(m)}(\Omega_i) = R_{ss}^{(m)}(\Omega_i) + R_{nn}^{(m)}(\Omega_i) + R_{bb}^{(m)}(\Omega_i).$$

Subtracting Eq. (18) from Eq. (17) and substituting with Eqs. (14) and (15) yields

$$\begin{aligned}&R_{yy}^{(m)}(\Omega_i) - R_{ee}^{(m)}(\Omega_i)\\ &\quad= R_{dd}^{(m)}(\Omega_i) - R_{bb}^{(m)}(\Omega_i)\\ &\quad= \frac{1-(F^{(m)}(\Omega_i))^2}{(1-F^{(m)}(\Omega_i))^2}R_{dd}^{(m)}(\Omega_i)\\ &\quad= \frac{1+F^{(m)}(\Omega_i)}{1-F^{(m)}(\Omega_i)}R_{dd}^{(m)}(\Omega_i), \quad F^{(m)}(\Omega_i) \neq 1\\ \Leftrightarrow\ &[1-F^{(m)}(\Omega_i)][R_{yy}^{(m)}(\Omega_i) - R_{ee}^{(m)}(\Omega_i)]\\ &\quad= [1+F^{(m)}(\Omega_i)]R_{dd}^{(m)}(\Omega_i).\end{aligned}$$

This leads directly to the result Eq. (19),

$$F^{(m)}(\Omega_i) = \frac{R_{yy}^{(m)}(\Omega_i) - R_{ee}^{(m)}(\Omega_i) - R_{\hat{d}\hat{d}}^{(m)}(\Omega_i)}{R_{yy}^{(m)}(\Omega_i) - R_{ee}^{(m)}(\Omega_i) + R_{\hat{d}\hat{d}}^{(m)}(\Omega_i)}.$$

Once again note that $F^{(m)}(\Omega_i) \neq 1$ means that there is some estimation of the echo at the frequency Ω_i, i.e. $R_{\hat{d}\hat{d}}^{(m)}(\Omega_i) \neq 0$ and $R_{yy}^{(m)}(\Omega_i) \neq R_{ee}^{(m)}(\Omega_i)$. Therefore, Eq. (19) is only valid for $R_{\hat{d}\hat{d}}^{(m)}(\Omega_i) \neq 0$.

References

[1] ITU-T Recommendation G.167, Acoustic Echo Controllers, Helsinki, 1993.

[2] Recommendation GSM 03.50, Transmission Planning Aspects of the Speech Service in the Public Land Mobile Network (PLMN) System, ETSI, Version 5.0.0, March 1996.

[3] H.W. Gierlich, The auditory perceived quality of hands-free telephones: Auditory judgements, instrumental measurements and their relationship, Speech Communication 20 (3–4) (1996) 241–254.

[4] S.F. Boll, Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. Acoust. Speech Signal Process. ASSP-27 (April 1979) 113–120.

[5] Y. Ephraim, D. Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process. 32 (6) (December 1984) 1109–1121.

[6] E. Hänsler, The hands-free telephone problem – An annotated bibliography, Signal Processing 27 (3) (1992) 259–271.

[7] E. Hänsler, The hands-free telephone problem – An annotated bibliography update, Ann. Télécomm. 49 (7–8) (1994) 360–367.

[8] Y. Ephraim, Statistical-model-based speech enhancement systems, Proc. IEEE 80 (10) (1992) 1526–1555.

[9] R. Wehrmann, R. Zelinski, Verfahren zur Geräuschreduktion bei der Sprachkommunikation über Freisprecheinrichtungen – eine Übersicht, Kleinheubacher Berichte, Band 38, Deutsche Telekom AG, 1995, pp. 483–495.

[10] J.R. Deller, J.G. Proakis, J.H.L. Hansen, Discrete-Time processing of Speech Signals, Macmillan, New York, 1993.

[11] H. Yasukawa, Acoustic echo canceller with sub-band noise cancelling, Electro. Lett. 28 (15) (1992) 1403–1404.

[12] Y. Grenier, M. Xu, J. Prado, D. Liebenguth, Real-time implementation of an acoustic antenna for audioconferencing, 1st Workshop on Acoustic Echo Control, Berlin, 1989.

[13] T. Huhn, H.-J. Jentschel, Kombination von Geräuschreduktion und Echokompensation beim Freisprechen, Nachrichtentechnik, Elektronik 43 (6) (1993) 274–280 (in German).

[14] R. Martin, P. Vary, Combined acoustic echo cancellation, dereverberation, and noise reduction: A two microphone approach, Proc. 3rd Internat. Workshop on Acoustic Echo Control, Lannion, 7–8 September 1993, pp. 125–132.

[15] R. Martin, P. Vary, Combined acoustic echo cancellation, dereverberation, and noise reduction: A two microphone approach, Ann. Télécomm. 49 (7–8) (1994) 429–438.

[16] B. Ayad, G. Faucon, Acoustic echo and noise cancelling for hands-free communication systems, Proc. 4th Internat. Workshop on Acoustic Echo and Noise Control, Roros, Norway, June 1995, pp. 91–94.

[17] G. Faucon, R. Le Bouquin Jeannes, Joint system for acoustic echo cancellation and noise reduction, Proc. EUROSPEECH'95, Madrid, 18–21 September 1995, pp. 1525–1528.

[18] R. Le Bouquin Jeannès, G. Faucon, B. Ayad, How to improve acoustic echo and noise cancellation using a single talk detector, Speech Communication 20 (3–4) (1996) 191–202.

[19] P. Scalart, A. Benamar, A system for speech enhancement in the context of hands-free radiotelephony with combined noise reduction and acoustic echo cancellation, Speech Communication 20 (3–4) (1996) 203–214.

[20] R. Martin, Combined acoustic echo cancellation, spectral echo shaping, and noise reduction, Proc. 4th Internat. Workshop on Acoustic Echo and Noise Control, Roros, Norway, June 1995, pp. 48–51.

[21] R. Martin, S. Gustafsson, The echo shaping approach to acoustic echo control, Speech Communication 20 (1996) 181–190.

[22] R. Martin, P. Vary, Combined acoustic echo control and noise reduction for hands-free telephony – State of the art and perspectives, Proc. EUSIPCO, 1996, Trieste, 1996, pp. 1107–1110.

[23] C. Antweiler, Orthogonalizing algorithms for digital compensation of acoustic echoes, in: P. Vary (Ed.), Dissertation, Aachener Beiträge zu Digitalen Nachrichtensystemen, Vol. 1, April 1995 (ISBN 3-86073-430-X) (in German).

[24] R. Frenzel, Freisprechen in gestörter Umgebung, Fortschritt-Berichte VDI, Reihe 10, VDI-Verlag, Düsseldorf, 1992.

[25] O. Cappé, Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor, IEEE Trans. Speech Audio Process. 2 (2) (1994) 345–349.

[26] R. Martin, Spectral subtraction based on minimum statistics, Proc. EUSIPCO-94, Edinburgh, 12–16 September 1994, pp. 1182–1185.

[27] G. Doblinger, Computationally efficient speech enhancement by spectral minima tracking in subbands, Proc. EUROSPEECH '95, Vol. 2, Madrid, 1995, pp. 1513–1516.

[28] R. Martin, Hands-free telephones based on multi-microphone echo cancellation and noise reduction, in: P. Vary (Ed.), Dissertation, Aachener Beiträge zu Digitalen Nachrichtensystemen, Vol. 1, June 1995 (ISBN 3-86073-432-6) (in German).

[29] S. Gustafsson, R. Martin, P. Vary, On the optimization of speech enhancement systems using instrumental measures, Proc. Workshop on Quality Assessment in Speech, Audio and Image Communication, Darmstadt, Germany, 11–13 March 1996.